(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(54) Title: ASSIGNMENT OF SEMANTIC TAGS TO PHRASES FOR GRAMMAR GENERATION
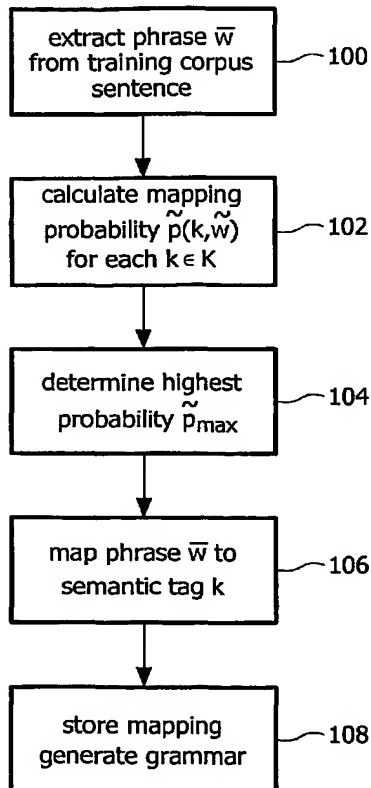
(57) Abstract: The present invention relates to a method, a system and a computer program product for mapping of semantic tags to phrases within a training corpus of weakly annotated sentences, thereby generating a grammar which can be applied to unknown sentences for the purpose of language understanding. The method is based on a probabilistic estimation that a given phrase is mapped to a semantic tag of a set of candidate semantic tags. The mapping and the generation of the grammar is performed according to a maximum mapping probability of a set of mapping probabilities of the given phrase and the set of candidate semantic tags. In particular, the determination of the mapping probability makes use of an expectation maximization algorithm.

extract phrase $\overline{w}$ from training corpus sentence — 100

calculate mapping probability $\tilde{p}(k,\overline{w})$ for each $k \in K$ — 102

determine highest probability $\tilde{p}_{max}$ — 104

map phrase $\overline{w}$ to semantic tag k — 106

store mapping generate grammar — 108

WO 2005/048240 A1

CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.